The Program for Climate Model Diagnosis and Intercomparison

Integrated and Distributed Software Solution for the Scientific Community
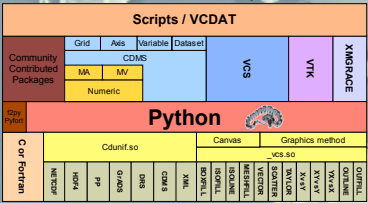
http://www-pcmdi.llnl.gov

PCMDI

Climate Data Analysis Tools
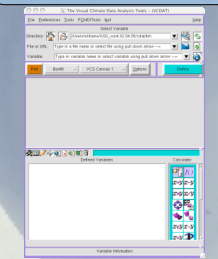

Climate Data Analysis Tools (CDAT)

The Climate Data Analysis Tools (CDAT) has addressed the challenges of enabling data management, discovery, access, and advanced data analysis of the important datasets that represent Earth system models. Well suited for gluing enterprise applications and component assembly (i.e., graphical user interfaces and web services), CDAT has become the next generation analysis tool for atmospheric scientists. Based on Python, a powerful user-friendly and widely used object-oriented scripting language, the computer and atmospheric scientists are provided with a significantly enhanced and productive working environment. At Lawrence Livermore National Laboratory, the Program for Climate Model Diagnosis and Intercomparison is central to the worldwide development and use of CDAT in climate and environmental studies. CDAT is extensively used by scientists to gain a better quantitative understanding of the Earth system and how its development is organized around a multi-layer community-based design, which promotes collaboration between expert computer scientists at its core and research scientists as the end-user.

CDAT Architectural Layers

CDAT

- **Python**: widely used dynamic object-oriented, scripting language with very simple syntax and clean semantics; easy to link with other languages (such as C and Fortran)
- **Climate Data Management System - CDMS**: provide access to and management of gridded data (such as NetCDF, GrADS/Grib, DHF, PP …)
- **Numerical Python**: large-array numerical operations
- **Visualization and Control System - VCS**: publication quality 1D and 2D graphics and animation
- **Contributed Modules**: community contributed climate diagnostic packages (such as R, spherepack, xmgrace, pyclimate, f2py, SCRIP, …)
- **PCMDI diagnostic modules**: genutil and cdutil contain such things as filters, statistics, interpolation, …)
- **Cdutil** (PCMDI developed) and a growing collection of community-based Packages: Atmospheric Chemistry, EOF, etc…)

VCDAT
Discover CDAT's feature and learn Python coding via a user-friendly GUI-based interface

Plone
Access Online Documentation, constantly updated and improved by the users community

Current CDAT Status

- Community-based Software: Ever-improving, ever-present
- Open-source Project, source code available freely, for community-based contributions
- 24/7 help from community, via mailing list and Plone website
- User-friendly GUI-based interface to major features of CDAT, let new users discover CDAT and learn how to code in Python
- Bugzilla: help developers track bugs and users see the status of their bugs
- Platforms: Linux, Unix, Mac OS X, MS Windows (via Cygwin)
- Integrated with other popular software components, such as: the popular Live Access Server (LAS), Open-source Project for a Network Data Access Protocol (OPeNDAP) and the Earth System Grid

Open Source
Code freely available, via free perforce client or subversion.

BUGZILLA
Web-based, bug tracking and reporting system.
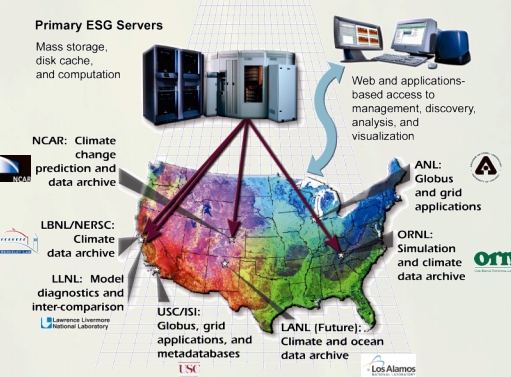

The Earth System Grid

Providing Climate Scientists with Virtual Proximity to the Large-Scale Simulation Results Needed for Their Research
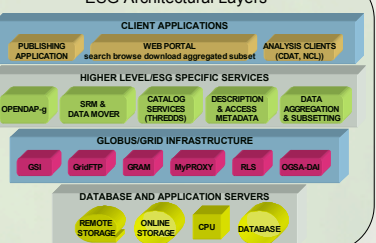
The Earth System Grid II (ESG) is a new research project sponsored by the U.S. DOE Office of Science under the auspices of the Scientific Discovery through Advanced Computing program (SciDAC). The primary goal of ESG is to address the formidable challenges associated with enabling analysis of and knowledge development from global Earth System models. Through a combination of Grid technologies and emerging community technology, distributed federations of supercomputers and large–scale data & analysis servers will provide a seamless and powerful environment that enables the next generation of climate research.

Primary ESG Servers

- **NCAR**: Climate change prediction and data archive
- **LBNL/NERSC**: Climate data archive
- **LLNL**: Model diagnostics and inter-comparison
- **USC/ISI**: Globus, grid applications, and metadatabases
- **ANL**: Globus and grid applications
- **ORNL**: Simulation and climate data archive
- **LANL (Future)**: Climate and ocean data archive

ESG Objectives

- Distributed data archives
- "Virtual datasets" created through subsetting and aggregation
- Simple web-based access with metadata-based search and discovery
- Efficient retrieval of numerous and/or large datasets
- ESG access via analysis clients, such as CDAT
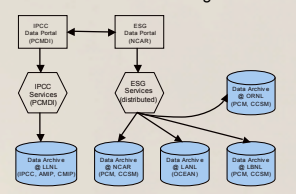
ESG Architectural Layers

Current ESG Status

- 885 datasets, comprising >30 TB of data in 60,000 files cataloged and available from the Intergovernmental Panel on Climate Change (IPCC) assessment runs; over 450 registered users
- 569 datasets, comprising >65 TB of data in 397,000 files cataloged and available from the Community Climate System Model (CCSM) and the Parallel Climate Model (PCM) runs; over 1230 registered users

ESG Data Holdings

## Serving Climate Scientific Community

IPCC Model Output

The Intergovernmental Panel on Climate Change (IPCC) was established by the World Meteorological Organization and the United Nations Environmental Program to assess scientific information on climate change. The IPCC publishes reports that summarize the state of the science. At the request of the JSC / CLIVAR Working Group on Coupled Modeling (WGCM), the PCMDI is archiving coupled ocean-atmosphere general circulation model output to support the Working Group 1 (WG1) component of the IPCC's 4th Assessment Report (AR4). The data archived by the PCMDI from each participating coupled ocean-atmosphere model is a subset of that model's output. Working Group 1 of the IPCC focuses on the physical climate system -- atmosphere, land surface, ocean and sea ice -- and the choice of variables archived at the PCMDI reflects this focus.

IPCC ESG Data Portal
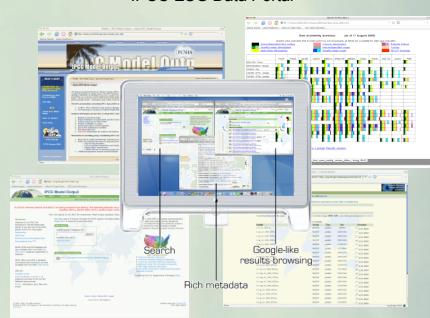
Figures for the IPCC WG1 AR4 Report
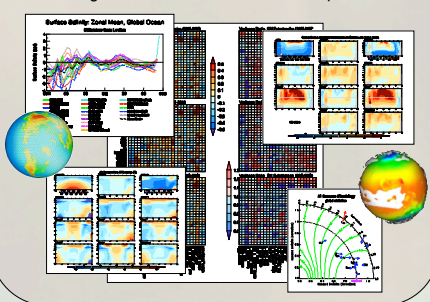
Data Discovery and Analysis

- PCMDI is hosting the IPCC WG1 AR4 data holdings
- CMOR libraries, developed at PCMDI, allow for outputting data in the standardized Climate Forecasting (CF) convention
- Scientists can early browse and download data needed for their research
- PCMDI's software suite, CDAT, allows for easy development of cutting-edge diagnosis & visualization, such as: "Taylor diagram", "Portrait Plot", easy access to multiple projection, 3D representation, . . .

Current IPCC Status

- Serving data to over 400 projects in the IPCC community
- Average data movement is between 300 and 400 gigabytes per day
- The IPCC archive contains over 30 terabytes from which over 200 research papers are actively being written

IPCC ESG services: information, metadata, community authorization

---